

## CLAIMS

What is claimed is:

1. A method for searching for a specific base sequence, comprising:

5        an acquisition step for a specific base sequence candidate, which acquires a specific base sequence candidate, which is a candidate of a specific base sequence appearing in a base sequence of an expressed gene;

         a searching step for a specific base sequence, which searches a matching base sequence, which is a base sequence matching the specific base sequence candidate acquired by said acquisition step for specific base sequence candidate, from a set of base sequences, which include  
10        a union of sets of

a union of sets of exon base sequences, and

a set of border base sequences, which straddle exon borders in the expressed gene formed by a plurality of exons; and

15        a determination step, which determines whether the specific base sequence candidate acquired by said acquisition step for a specific base sequence candidate is a specific base sequence based on whether a plurality of matching base sequences are included in the search result by said search step for a specific base sequence.

20

2. The method for searching for a specific base sequence according to Claim 1, wherein

attribute information including information indicating the position of exon sequence, or information for identifying gene formed by exon, is correlated to an element of said union of set of exon base sequences.

5           3. The method for searching for a specific base sequence according to Claim 1 or 2, wherein

          said set of border base sequences is acquired based on a set acquired by integrating information indicating a base sequence, which has same expressed gene and overlapping position of base sequence, to the set of information, which indicates a base sequence straddling the exon  
10   border in the expressed gene formed by a plurality of exons, and indicates the base sequence of the same length as that of the base sequence of said specific base sequence candidate.

          4. The method for searching for a specific base sequence according to any one of Claims 1 to 3, comprising:

15           an acquisition step for allowable number of matches, which acquires a numerical value, indicating the number of allowable mismatching bases, as a degree of matching between the base sequence included in said set of base sequences and the base sequence indicated by said specific base sequence candidate, wherein

          said searching step for base sequence carries out search based on the allowable number of  
20   matches acquired by said acquisition step for allowable number of matches.

5. The method for searching for a specific base sequence according to Claim 4, comprising:

an acquisition step for mismatching base pair, which acquires a base pair, which is determined to be mismatching by said searching step for base sequence.

5 6. The method for searching for a specific base sequence according to any one of Claims 1 to 5, comprising:

an acquisition step for distribution information of mismatching, which acquires distribution information indicating a distribution of occurrence of mismatching bases as a degree of matching between the base sequence included in said set of base sequences and the base sequence indicated

10 by said specific base sequence candidate, wherein

said searching step for base sequence carries out search based on the distribution information acquired by said acquisition step for distribution information of mismatching..

7. The method for searching for specific base sequence according to Claim 6, wherein

15 said distribution information indicates length of successive bases, which are not determined to be mismatching.

8. The method for searching for specific base sequence according to any one of Claims 1 to

20 7, wherein

said specific base sequence candidate is a candidate of a base sequence of oligo-DNA for

microarray.

9. The method for searching for a specific base sequence according to any one of Claims 1 to 7, wherein

5        said specific base sequence candidate is a candidate of base sequence of siRNA.

10. An apparatus for searching for a specific base sequence, comprising:

a storage for set of base sequences, which stores a set of base sequences, which includes a union of sets of

10        a union of sets of exon base sequences, and

a set of border base sequences, which straddles exon border in the expressed gene formed by a plurality of exons;

an acquirer for specific base sequence candidate, which acquires a specific base sequence candidate, which is a candidate of a specific base sequence appearing in a base sequence of an  
15        expressed gene; and

a searcher for specific base sequence, which searches for a matching base sequence, which is a base sequence matching the specific base sequence candidate acquired by said acquirer for specific base sequence candidate, from the base sequences included in the set of base sequences stored by said storage for set of base sequences.

20

11. The apparatus for searching for specific base sequence according to Claim 10, wherein

attribute information, including information indicating position of exon sequence, or information for identifying gene formed by exon, is correlated with an element of said union of sets of exon base sequences.

5           12. The apparatus for searching for a specific base sequence according to Claim 10 or 11, wherein

said set of border base sequences is acquired based on a set acquired by integrating information indicating a base sequence, which has the same expressed gene and overlapping position of base sequence, as the set of information, which indicates a base sequence straddling the  
10   exon border in the expressed gene formed by a plurality of exons, and indicates the base sequence of the same length as that of the base sequence of said specific base sequence candidate.

13. The apparatus for searching for specific base sequence according to any one of Claims 10 to 12, comprising:

15           an acquirer for allowable number of matches, which acquires a numerical value, indicating the number of allowable mismatching bases, as a degree of matching between the base sequence included in said set of base sequences and the base sequence indicated by said specific base sequence candidate, wherein

said searcher for base sequence carries out search based on the allowable number of  
20   matches acquired by said acquirer for allowable number of matches.

14. The apparatus for searching for a specific base sequence according to Claim 13,  
comprising:

an acquirer for mismatching base pair, which acquires a base pair, which is determined to  
be mismatching by said searcher for base sequence.

5

15. The apparatus for searching for a specific base sequence according to any one of  
Claims 10 to 14, comprising:

an acquirer for distribution information of mismatching, which acquires distribution  
information indicating a distribution of occurrence of mismatching bases as degree of matching  
between the base sequence of said set of base sequence and the base sequence of said specific base  
sequence candidate, wherein

10

said searcher for base sequence carries out search based on the distribution information  
acquired by said acquirer for distribution information of mismatching.

15

16. The apparatus for searching for a specific base sequence according to Claim 15,  
wherein

said distribution information indicates length of successive bases, which are not determined  
to be mismatching.

20

17. An apparatus for storing set of base sequences, storing

a set of base sequences including

a union of sets of exon base sequences, and

a set of border base sequences straddling exon border in the expressed gene formed by a plurality of exons, in a searchable state.

5

18. The apparatus for storing a set of base sequences according to Claim 17, wherein attribute information, including information indicating position of exon sequence, or information for identifying gene formed by exon, is correlated to an element of said union of sets of exon base sequences.

10

19. The storage for set of base sequence according to Claim 17 or 18, wherein said set of border base sequences is acquired based on a set acquired by integrating information indicating a base sequence, which has the same expressed gene and overlapping position of base sequence, to the set of information, which indicates a base sequence straddling the exon border in the expressed gene formed by a plurality of exons, and indicates the base sequence of the same length as that of the base sequence as an input for searching.

15

20. A generation method for set of base sequence, comprising:

an acquisition step for length of base sequence candidate, which acquires length of specific

20 base sequence candidate appearing in a base sequence of an expressed gene;

an acquisition step for set of exon base sequences, which acquires a union of sets of exon

base sequences;

a generation step for set of border base sequences, which generates a set of base sequences by integrating information indicating a base sequence, which has the same expressed gene and overlapping position of base sequence, to the set of information, which indicates a base sequence straddling the exon border in the expressed gene formed by a plurality of exons, and indicates the base sequence of the same length as that acquired by said acquisition step for length of base sequence candidate; and

a generation step for union of sets, which generates a union of sets of the base sequences acquired by said acquisition step for set of exon base sequences, and set of the base sequences generated by said generation step for set of border base sequences.

21. A searching program for specific base sequence, causing a computer to carry out:

an acquisition step for specific base sequence candidate, which acquires a specific base sequence candidate, which is a candidate of a specific base sequence appearing in a base sequence of an expressed gene; and

a search step for a specific base sequence, which searches for a matching base sequence, which is a base sequence matching a base sequence indicated by the specific base sequence candidate acquired by said acquisition step for a specific base sequence candidate, from a set of base sequences, which includes a union of sets of

a union of sets of exon base sequences, and

a set of border base sequences, which straddles exon borders in the expressed gene



formed by a plurality of exons.

22. A generation program for a specific base sequence, causing a computer to carry out:

an acquisition step for length of base sequence candidate, which acquires length of specific

5 base sequence candidate appearing in a base sequence of an expressed gene;

an acquisition step for set of exon base sequences, which acquires a union of sets of exon  
base sequences;

a generation step for set of border base sequences, which generates a set of base sequence  
by integrating information indicating a base sequence, which has same expressed gene and

10 overlapping position of base sequence, to the set of information, which indicates a base sequence  
straddling the exon border in the expressed gene formed by a plurality of exons, and indicates the  
base sequence of the same length as that acquired by said acquisition step for length of base  
sequence candidate; and

a generation step for union of sets, which generates a union of set of the base sequences

15 acquired by said acquisition step for set of exon base sequences, and set of the base sequences  
generated by said generation step for set of border base sequences

23. A search program for a specific base sequence, causing a computer to carry out:

an acquisition step for a specific base sequence candidate, which acquires a specific base

20 sequence candidate, which is a candidate of a specific base sequence appearing in a base sequence  
of an expressed gene;

a search step for a specific base sequence, which searches for a matching base sequence, which is a base sequence matching a base sequence indicated by the specific base sequence candidate acquired by said acquisition step for specific base sequence candidate, from a set of base sequences, which includes a union of sets of

5 a union of sets of exon base sequences, and

a set of border base sequences, which straddles exon border in the expressed gene formed by a plurality of exons; and

a determination step, which determines whether the specific base sequence candidate acquired by said acquisition step for specific base sequence candidate is a specific base sequence  
10 based on whether a plurality of matching base sequences are included in the search result by said searching step for specific base sequence.